Design of Communication Processor Chips with New Shared Memory(新しい共有キャッシュメモリを有するコミュニケーションプロセッサの設計)

著者	張 俊
号	3101
発行年	2003
URL	http://hdl.handle.net/10097/8373

氏名	ジャン ジュン 張 俊
授与学位	博士（工学）
学位授与年月日	平成16年3月25日
学位授与の根拠法規	学位規則第4条第1項
研究科，専攻の名称	東北大学大学院工学研究科（博士課程）機械知能工学専攻
学位論文題目	Design of Communication Processor Chips with New Shared Cache Memory （新しい共有キャッシュメモリを有するコミュニケーションプロセッサの設計）
指導教官	東北大学教授 小柳 光正
論文審査委員	主査 東北大学教授 小柳 光正 東北大学教授 羽根 一博 東北大学教授 中村 維男 東北大学助教授 栗野 浩之 （情報科学研究科）

論文内容要旨

Chapter 1 Introduction

The network traffic in today's networks is growing at an astonishing rate. It has become the main challenge to the Internet service providers to offer the effective capability of supporting increasing network traffic. Eliminating network bottlenecks continues to be a top priority for the Internet service providers. Routers are often the source of these bottlenecks. Since the carriers that provide the Internet's backbone networks have rapidly added bandwidth, more focus is placed on new network devices that can utilize bandwidth resources more efficiently and provide the advanced data services at wire speed, that are commonly found in routers and network application servers. Moreover, such new network devices also need enough flexibility to support new protocols and applications. The purpose of this work is just to offer an alternative scheme, called communication processors, to satisfy the requirements of intelligent network processing with both high performance and good flexibility.

A communication processor can be defined as a programmable communications integrated circuit with special hardware structure for packet processing. It is an ASIP (Application Specific Instruction Processor) capable of performing the underlying network functions such as lookup, data parsing and classification, computation, data transformation, media access control (MAC), queue management and control processing.

Chapter 2 Chip Architecture and Routing Operations

The designing of communication processors is driven by the requirements of both performance and flexibility. RISC-based chip architecture is employed to ensure high-level programmability for emerging network application requirements. Moreover, new IP routing lookup algorithm and special hardware solutions, such as special Table Lookup Unit, parallel processing, pipelined processing, shared data cache and special bus architecture, are used to provide speed improvement for basic network processing functions.

A new chip architecture is proposed in Fig. 1. Four 32-bit RISC processor elements (PEs) are embedded on the chip and can be configured to a parallel processing system block to provide high-performance computation and low latency. A new configurable-shared data cache is proposed to offer immediate data sharing for four RISC PEs. A special Table Lookup Unit (TLU) can implement high-performance IP routing lookups and updates, which offloads the complex, time-consumed lookup tasks from software on the RISC PEs. Four Ethernet MAC Units (EthMAC) can implement IEEE 802.3 Medium Access Control functions and support 100-Mb/s or 10-Mb/s data rates. DMA Controller is responsible of

handling low-latency direct memory access from EthMACs or from HC Interface and offering independent memory-to-memory data movement. SDRAM Controller is responsible of controlling up to 256Mbytes SDRAM and performing SDRAM burst read or write accesss. Queue Management Unit queues transmitting descriptors for output packets. Bus Arbiter arbitrates between multiple simultaneous requests for the controlling of the configuration bus. HC Interface provides an external DMA interface to construct a larger communication system with higher processing power. Instruction Cache provides instruction fetching for RISC PEs. All the units are connected together by two on-chip buses in a communication processor. The data-specific bus offers memory-to-memory DMA transfers, transmitting or receiving DMA transfers between SDRAM memory and EthMACs or HC Interface. The configuration bus offers program control to Table Lookup Unit, Queue Management Unit, DMA Controller, Bus Arbiter and HC Interface.

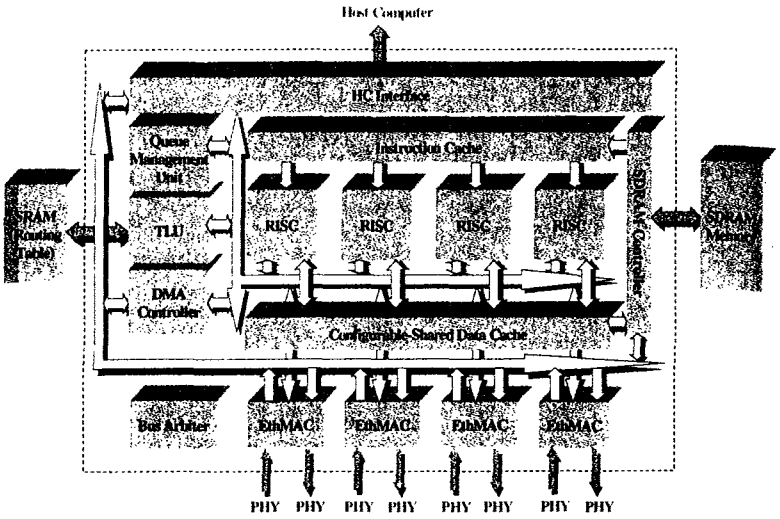

Fig. 1 Chip architecture of communication processors

The architecture features and routing operations of communication processors are also described in this chapter.

Chapter 3 Design and Evaluation of New High Speed IP Routing Lookup Structure

The exponential growth of the Internet has burdened the IP routing lookup system that plays an important role in packet's hop-by-hop forwarding path. While the link speed continues to increase, the IP routing lookup has become a major bottleneck in forwarding performance of communication processors. Therefore, a special Table Lookup Unit with effective IP routing lookup structure can relieve load from the RISC PEs so that they can provide maximum computational throughput for the processing of higher layer protocols. This chapter describes a pipelined hardware structure with selective binary search algorithm for Table Lookup Unit.

At first Chapter 3 presents a survey of some latest IP routing lookup schemes that can be classified approximately into four classes: trie-based scheme, CAM-based scheme, cache-based scheme and binary-search-based scheme. The drawbacks of these existing schemes are also discussed.

Based on basic binary search on prefix lengths, a new IP routing lookup algorithm, called selective binary search algorithm, is proposed to avoid the unacceptable update overhead generated by precomputation. By replacing precomputation with several NPM fields indicating the number of the longer prefixes that can be matched, we reduce the search steps required by one lookup greatly even if the backtracking happens. A new pipelined hardware structure including 24 sub-search units is also proposed to implement the selective binary search. We use caching to reduce the hash operations required by each sub-search unit greatly, because we believe that the first few bits have greater locality than the individual IP address itself. Moreover, in contrast with generic cache in computer, cache adopting nonexistent prefixes reduces the miss ratio effectively. Because caching makes each sub-search unit run independently, pipelining can be used for the whole system and a maximum rate of one lookup per cycle is possible.

The performance evaluation of Table Lookup Unit shows that the hash operation ratio for one lookup is reduced to less than 1% and less than two hash operations are needed for one update. Moreover, only 512Kbytes are needed for our sample routing table with about 43000 prefixes. The overhead for lookup, update and storage proves that our scheme have better performance than the existed schemes.

At last, the new scheme can be also expanded to the IPv6 routing table lookup problem. A three-dimensional structure is also discussed. Future work on our scheme will be focused on designing a three-dimensional structure to solve the serious interconnection problem by using a new three-dimensional

integration technology.

Chapter 4 Design and Evaluation of Parallel Processing System Block with New Configurable-Shared Cache Structure

Since packet-processing applications have high parallelism in computations, four RISC PEs are embedded on a communication processor chip to construct a parallel processing system block and offer outstanding computational throughput and low interactive response time. The RISC PE structure and instruction set are introduced in section 4.2. The working modes of the parallel processing system block is introduced in section 4.3.

How to implement data sharing among different RISC PEs is a very important point in the designing of such on-chip parallel processing system block. Sharing on cache level is an effective alternative not only to reduce the communication load on outside system, but also to narrow the performance gap between the chip and external memory. Section 4.4 presents a survey of several existing shared cache schemes including conventional multiport SRAM structure, snoop cache structure and pSAS (Pseudo Set Associative and Shared) cache structure. The drawbacks are also discussed respectively.

A new shared cache structure, called configurable-shared data cache structure, is proposed in Fig. 2. Each RISC PE has a private generic cache block. All the cache blocks are also connected by an on-chip broadcast bus. Each PE can access the respective cache block simultaneously and independently from other PEs. The broadcast controller decides which PE is permitted to use the broadcast bus as well as when the broadcast bus is used. The refill/writeback controller controls the activity of refill or writeback operations. According to the configuration to broadcast controller, two kinds of work modes are available in the configurable-shared data cache: a shared-cache mode for the applications with fine-grained parallelism and high degree of data sharing, and a multi-program mode for multi-program applications.

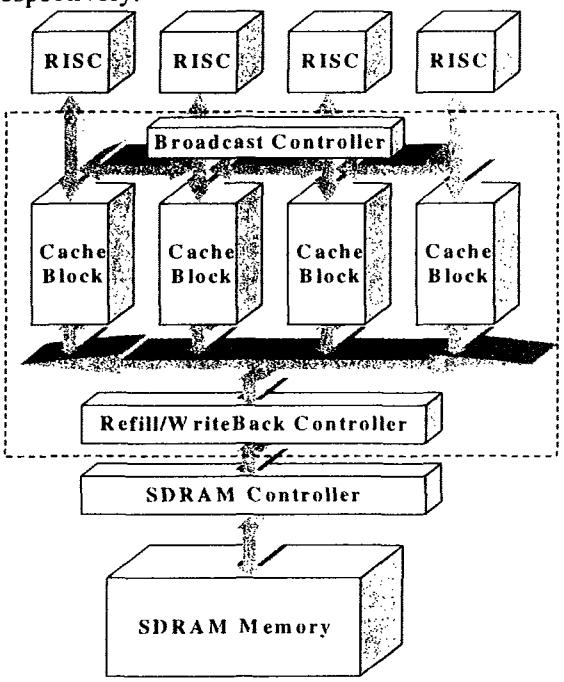

Fig. 2 Configurable-shared cache structure

The evaluation result shows that the performance of parallel processing system block with configurable-shared data cache structure improves in proportion to the number of RISC PEs. Though the effective utilization of cache block decreases comparatively with the PE number, it is profitable to trade large cache for the high performance because generic cache is used. Moreover, the configurable-shared data cache structure has the advantage in scalability, and there is no worry that the access time increases with the port number, or the chip area increases more than the linear proportion of the port number.

To solve the problem in the complexity of wiring when constructing a large-scale configurable-shared data cache structure, chapter 4 also proposes a consideration about a three-dimensional configuration (3D-CSHDC) by using three-dimensional integration technology.

Chapter 5 Chip Implementation and Performance Evaluation of Communication Processor

The whole chip is implemented on RTL mode in VHDL. A bottom-up design methodology is adopted. Since Table Lookup Unit and the parallel processing system block consisting of four RISC and a configurable-sharing data cache have been described in chapter 3 and chapter 4 respectively, chapter 5 describes the implementation of other main function units: DMA Controller, Ethernet MAC Unit, SDRAM Controller, Queue Management Unit, Bus Arbiter and HC interface. Three low-latency DMA transfers are described together with each function units. Table Requests to Table Lookup Unit is also described. Moreover, a large communication system with higher processing power and throughput constructed by using

multiple communication processors is also discussed.

The evaluation result shows that the routing latency decreases to 44% of that in a single RISC routing system, while the latency of IP packet processing decreases to 12% of that in a single RISC routing system. Therefore, the new architecture of communication processors can provide performance improvement greatly for network devices as well as ensure enough programmability. The evaluation also shows the packet processing power of communication processors is not utilized efficiently if it is employed to 100Base Ethernet network. Therefore, later work will be focused on implementing EthMAC to higher bandwidth such as Gigabit EthMAC (1000Base) and evaluating the performance.

Chapter 6 Conclusion

In conclusion, this work proposes a new chip architecture for communication processors. The new architecture is capable of offering high-performance processing power as well as high-level programmability for network applications. The communication processors can also be used to construct a large communication system with higher processing power and throughput.

論文審査結果の要旨

インターネットの急速な普及とネットワーク規模の拡大に伴って、高速でかつ高信頼性のネットワーク経路探索機能をもつ新しいコミュニケーションプロセッサ・チップに対する要求が増している。しかし、従来のバイナリーサーチ・アルゴリズムに基づいた経路探索機能とそれを実現するためのハードウェア技術を用いる手法では、経路探索に必要な計算のオーバーヘッドや処理のレーテンシィのために、高速でかつ高信頼性のネットワーク経路探索機能をもつコミュニケーションプロセッサ・チップを実現することが難しい。本論文は、新しいサーチアルゴリズムとそれを実行するためのパイプライン処理、さらにはリコンフィグュラブル共有キャッシュメモリと並列処理動作を導入することによって、高速でかつ高信頼性のネットワーク経路探索機能をもつコミュニケーションプロセッサ・チップを実現できることを明らかにしたもので、全編６章よりなる。

第１章は緒言である。

第２章では、４個のRISC プロセッサと共有キャッシュメモリからなる並列処理ブロックと２種類のデータバスを有する新しいコミュニケーションプロセッサ・チップを提案し、性能改善や低消費電力化に対する効果およびソフトウェア対応の柔軟性について考察している。これらは有用な知見である。

第３章では、ＩＰネットワーク経路探索におけるアドレス計算のオーバーヘッドを軽減するための新しい手法である選択的バイナリーサーチ・アルゴリズムを提案するとともに、これを高速に実行するためのパイプライン方式を提案している。また、サブサーチユニットのハッシュ動作にキャッシュを用いた新しいテーブルルックアップユニットを導入することによって１回のルックアップに必要なハッシュ動作比率を１％以下にするとともに、１回のアドレス更新に必要なハッシュ動作を２回以下に抑えることができることをシミュレーションにより確認している。これらの成果は極めて重要で、高く評価できる。

第４章では、コミュニケーションプロセッサ・チップ内部の並列処理ブロックの詳細設計について述べている。この並列処理ブロックでは、共有キャッシュモードとマルチプログラムモードという２つの動作モードをもつ新しいリコンフィグュラブル共有キャッシュメモリが使われており、これによってレーテンシィやスケーラビリティを改善し高性能化を実現している。これは有用な成果である。

第５章では、テーブルルックアップユニット、並列処理ブロックにDMAコントローラ、イーサーネットMACユニット、ＳＤＲＡＭコントローラ等を加えたコミュニケーションプロセッサ・チップ全体の詳細設計を行い、シミュレーションによりチップ性能を評価している。その結果、１個の RISC プロセッサを用いた場合に比べて、経路探索のレーテンシィを４４％に、またＩＰパケット処理のレーテンシィを１２％に低減できることを明らかにしている。これらは有用な成果である。

第６章は結言である。

以上、要するに本論文は、新しいサーチアルゴリズムとそれを実行するためのパイプライン処理、さらにはリコンフィグュラブル共有キャッシュメモリと並列処理動作の導入によって、高速でかつ高信頼性のネットワーク経路探索機能をもつコミュニケーションプロセッサ・チップを実現できることを明らかにしたもので、機械知能工学の発展に寄与するところが少なくない。

よって，本論文は博士(工学)の学位論文として合格と認める。